Developing Enculturated Agents

Pitfalls and Strategies

Rehm, Matthias

Published in:
Handbook of Research on Culturally-Aware Information Technology

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Handbook of Research on Culturally-Aware Information Technology:
Perspectives and Models

Emmanuel G. Blanchard
McGill University, Canada

Danièle Allard
Université de Sherbrooke, Canada

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE

Hershey • New York

Director of Editorial Content: Kristin Klinger
Director of Book Publications: Julia Mosemann
Acquisitions Editor: Lindsay Johnston
Development Editor: Christine Buffton
Publishing Assistant: Travis Gundrum and Natalie Pronio
Typesetter: Deanna Jo Zombro
Production Editor: Jamie Snavely
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on culturally-aware information technology : perspectives and models / Emmanuel Blanchard and Daniele Allard, editors.
p. cm.
Includes bibliographical references and index. Summary: "This book provides readers with the possibility of acquiring in-depth knowledge of the theoretical and technological research conducted in IT in relation to culture"--Provided by publisher.
ISBN 978-1-61520-883-8 (hbk.) -- ISBN 978-1-61520-884-5 (ebook) 1. Information technology--Social aspects. 2. Information technology--Cross-cultural studies. I. Blanchard, Emmanuel (Emmanuel G.) II. Allard, Daniele, 1964-
HM851.H3482 2010
303.48'33--dc22
2009053457

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 16
Developing Enculturated Agents:
Pitfalls and Strategies

Matthias Rehm
Aalborg University, Denmark

ABSTRACT

Embodied Conversational Agents (ECAs) are complex multimodal systems with rich verbal and nonverbal repertoires. There human-like appearance raises severe expectations regarding natural communicative behaviors on the side of the user. But what is regarded as "natural" is to a large degree dependent on our cultural profiles that provide us with heuristics of behavior and interpretation. Thus, integrating cultural aspects of communicative behaviors in virtual agents and thus enculturating such systems seems to be inevitable. But culture is a multi-defined domain and thus a number of pitfalls arise that have to be avoided in the endeavor. This chapter presents some of the pitfalls for enculturating interactive systems and presents strategies on how to avoid these pitfalls in relation to the standard development process of Embodied Conversational Agents.

ENCULTURATED AGENTS: A DEFINITION

This chapter argues that Embodied Conversational Agents (ECAs) (Cassell, Sullivan, Prevost & Churchill, 2000) are prototypical devices for enculturating the human computer interface. It examines the standard development process for ECA systems and discusses at each step strategies to avoid the pitfalls that arise from integrating culture as a computational parameter into the process. The chapter is not going to argue for or against specific cultural theories, but relies on Hofstede's (2001) dimensional theory of culture as a widely used example.

Embodied Conversational Agents can be regarded as a special case of multimodal dynamic interactive systems (see Figure 1 for some examples). They promote the idea that humans, rather than interacting with tools prefer to interact with an artifact that possesses some human-like qualities. If it is true, as Reeves and Nass' (1996)

DOI: 10.4018/978-1-61520-883-8.ch016

media equation suggests, that people respond to computers as if they were humans, then there are good chances that people are also willing to form social relationships with virtual agents. As a consequence, it seems inevitable to take cultural aspects into account when creating such agents. Due to their embodiment, agents present complex multimodal systems with rich verbal and nonverbal repertoires. Additionally, the appearance of the agent might play an important role when taking cultural aspects into account.

Embodied Conversational Agents as an interface metaphor have a great potential to realize cultural aspects of behavior in several fields of human computer interaction:

1. Information presentation: By adapting their communication style to the culturally dominant persuasion strategy, agents become more efficient in delivering information or selling a point or a product.
2. Entertainment: Endowing characters in games with their own cultural background has two advantages. It makes the game more entertaining i.) by providing coherent behavior modifications based on the cultural background and ii.) by letting the characters react in a believable way to (for them) weird behavior of other agents and the user.
3. Education: For educational purposes, experience-based *role-plays* become possible, e.g. for increasing cultural awareness of users or for augmenting the standard language textbook with behavioral learning.

The following issues for enculturating Embodied Conversational Agents are discussed in this chapter:

1. Enculturating agents opens up a challenging research field because culture penetrates most of the above mentioned features (verbal and nonverbal behavior, appearance) of an agent. Thus, enculturating such a system has to rely on a solid theoretical framework that is able to describe or even predict these influences.
2. Another critical issue that has to be discussed but is not easily solved is the following: apart from a specific cultural theory, different levels of culture like national culture, regional culture, the culture of the agents

Figure 1. Examples of Embodied Conversational Agents. Top row: affective spectators (Damian, Janowski & Sollfrank, 2009), an autonomous bot in second life (Rehm & Rosina, 2008), the Gamble multiuser dice game (Rehm, 2008), interacting with virtual dancers (Rehm, Vogt, Bee & Wissner, 2008). Bottom row: collaborating agents in edutainment (Rehm, André, Conradi, Hammer, Iversen, Lösch, Pajonk & Stamm, 2006), a virtual tourist guide, the FearNot! anti-bullying system (Hall, Woods, Aylett, Newall & Paiva, 2005).

vs. the culture of the developer have to be regarded.

3. Moreover, the developers' own cultural background provides them with implicit design heuristics for the system, which have to be challenged actively at every step of the process.

These issues are addressed in two ways. By reviewing the current research landscape for cultural aspects of ECA systems it becomes clear that the importance of cultural variations in interaction have been acknowledged but that there is a struggle for reliable models and representations of cultural aspects. With this knowledge at hand, pitfalls and strategies are discussed in relation to the methodological approach for realizing ECA systems, the study-model-test development process. The consideration of culture-specific interactions raises tough challenges on all levels of this process.

The chapter starts with identifying the pitfalls of enculturating ECAs on a theoretical and a practical level. A review of state of the art systems that claim to incorporate cultural aspects of interaction then exemplifies if and how the issues raised are currently addressed. Afterwards, strategies for enculturating ECA systems are presented in relation to the standard development process.

IDENTIFYING THE PITFALLS: THEORETICAL AND PRACTICAL ISSUES OF INTEGRATING CULTURAL ASPECTS INTO INTERACTIONS WITH CONVERSATIONAL AGENTS

Culture is a quite fuzzy notion and constitutes a multi-defined domain in the sense that multiple conflicting theories exist. Thus, cultural influences on the design process as well as on the interaction with the resulting system are not easy to define. In this section, we raise a number of questions that are used in the remainder of the chapter to structure the review of existing systems and the design process of ECA systems itself. Some of these questions are of a fundamental and theoretical nature others deal with practical issues that arise in the development of culturally aware interactive systems. The issues raised here are inspired by Blanchard and Mizoguchi's (2008) identification of culture-related topics in the development of culturally aware tutoring systems, which is one of the application domains of enculturated conversational agents.

Theoretical Issues

To tackle the challenge of enculturating complex interactive systems we have to start with the most fundamental question which is concerned with the

Definition of Culture

From the numerous approaches that define culture, we can focus our attention on those theories or concepts of culture that feature a way to describe, predict and measure the impact of culture on observable behavior. This allows us to utilize the notion of culture as a computational parameter that can be integrated into our systems. As will be apparent from the system reviews in the next section, the predominant cultural model is Hofstede's (2001) dimensional approach and Hall's (1959) dichotomies regarding context, contact, and sequencing of actions. Both theories are part of a school of thinking that defines culture as a commonly shared set of norms or values that influence behavior. Other proponents of this line are for instance Kluckhohn & Strodtbeck (1961) or Schwartz & Sagiv (1995). A more recent survey of cultural differences in organizations from 62 societies is the GLOBE study by House and colleagues (2004). All these approaches have in common that they define culture by focusing on differences in behavior. This view has been challenged by another line of research, which

argues that individuals do not belong to monolithic cultures but are embedded in a multitude of context-dependent social activities that all contribute to an individual's identity. It remains to be shown, how such an approach can be adapted in a computational system because it means the system has to answer the question

How to Cope with the Multiplicity of the Concept of Culture

The above mentioned theories focus generally on national cultures, which is a good starting point as it keeps some of the complexity at check. But actually, people are embedded in layers of cultural contexts, i.e. in a multitude of groups among which values and norms are shared, e.g. music culture (Pop vs. Metal vs. Gothic), sports culture (golf vs. swimming), professional culture (bus driver vs. bank accountant), national culture, religion, age, gender, etc. By reducing culture to national cultures, it becomes feasible to define behavioral heuristics for ECA systems that depict some broadly accepted standards, but the real challenge is dealing appropriately with the different layers of cultural profiles. An additional problem is to prevent the systems from just replicating stereotypes of cultures. A feasible approach seems to be to rely on general culturally determined heuristics of behavior that are overlayed by idiosyncratic behavior of a specific agent, for instance defined by its personality traits, its individual interaction history or its current emotional state.

How to Cope with Different Degrees of Cultural Heuristics

Individuals tend to exhibit a range of different idiosyncratic behaviors. Thus, even if a cultural model is able to describe or predict patterns of behavior, these are not necessarily being seen in every individual of the target group at every time and thus the model might not be applicable to predict the specific behavior of a given individual in a particular situation. This is not surprising as culture is primarily a social group phenomenon. Thus, in a given culture, cultural patterns of behavior are seen as common heuristics on how to behave "properly" and on how to interpret behaviors of others. As with all heuristics, the actual behavior in a specific situation is also dependent on a multitude of other contextual factors like emotional state, age of the interlocutor, social status, and location where the interaction is taking place, and many more.

Practical Issues

Apart from the theoretical issues, developing culturally aware interactive systems raises some quite practical questions on how to integrate culture into the system. First of all, a developer needs to decide

Which Layers of Cultural Influence Should be Addressed

Due to its fuzzy nature, culture penetrates most processing steps of an ECA system starting with superficial features like the agent's appearance (dress, skin color, etc.) over the agent's verbal and nonverbal behavior like the language that is used in the system, the dialog strategies that are employed, the gestural activity of the agent or its gaze behavior. On a cognitive level, cultural heuristics could also influence the structuring and use of knowledge bases and the appraisal of incoming information. On a more abstract level, task design and decomposition is also affected by culture for instance when defining the curriculum for an intelligent tutoring system (Hayashi, Bourdeau & Mizoguchi, 2008).

How to Prevent a Bias by the Developer's Own Cultural Background

The cultural background of the developer itself might be a problem during the process because it implicitly suggests design guidelines based on the cultural heuristics. For instance, when

creating animations for conversational gestures, the designer's cultural heuristics might interfere by providing him with guidelines on the spatial extent of a gesture which might not be appropriate for the target culture. The next section gives examples on how this can result in problems for the interaction with the system as well as examples on how this bias can be changed into a feature for the development of the system.

How to Get Reliable Data

Having identified the aspects that should be integrated in the system (for instance gaze and turn-taking behavior), the next challenge is to find the data about cultural heuristics that are relevant for these aspects. A literature review is a good starting point but some have argued (e.g. Jan, Herrera, Martinovski, Novick & Traum, 2007; Rehm, André, Bee, Endrass, Wissner, Nakano, Lipi, Nishida & Huang, 2009) that the information found in the literature is often too unspecific or incomplete to be useful for the development of interactive systems. Observational studies seem to be an alternative but raise other challenges on ensuring the quality of the data.

How to Cope with the Multiplicity of Influences on Behavior

Generally speaking, culture can be seen as a contextual factor influencing the interaction with a complex system. But culture is not the only contextual factor. Others include personality or emotional state, and many more. Thus, it has to be shown how this multiplicity of contextual factors can be dealt with to create believable and consistent behavior for an ECA system.

How to Assess the Results of Integrating Cultural Aspects

The multiplicity of influencing factors poses another challenge. If a system claims to have integrated cultural aspects of behavior, how can the success of this integration be measured. Above it was argued that culture penetrates all processing steps of an agent system. Thus, if a system integrates a specific aspect of culture-specific behavior like gestural activity but does not address facial expressions, is it possible to measure the effect of the culturally adequate gestural behavior?

We do not claim that the above mentioned issues are in any way fully inclusive. But we regard them as fundamental to ensure the success of an enculturated system. Consequently, these issues are possible pitfalls for realizing enculturated conversational agents. The rest of the chapter is dedicated to the quest of pointing out strategies to prevent these pitfalls. First, existing systems are reviewed in the light of the issues raised here and the strategies used in these systems are highlighted. Then, a general methodological approach of designing ECA systems is introduced and an analysis of options is presented for each step in the development process.

REVIEW OF CULTURE-AWARE SYSTEMS

The review focuses mainly on systems that incorporate Embodied Conversational Agents as their primary interface metaphor but in order to allow for a more general assessment of strategies in culturally aware systems, it also includes some agent-free systems that point out relevant options for circumventing the pitfalls. The order in which the work is presented stems from if and how they address the above mentioned issues, starting with work that is explicitly based on specific cultural theories and ending with work that mainly focuses on technical aspects. Tables 1,2 and 3 give an overview of this analysis.

Blanchard and Mizoguchi (2008) work in the area of culturally aware tutoring systems and propose the use of a high level structuring device in the form of upper ontologies to challenge the multi-defined domain of culture. Their upper-ontology of culture represents the concepts and

Table 1. Summary of how the systems reviewed in this section address the identified pitfalls.

	Definition of culture	**How to cope with the multiplicity of the concept of culture?**	**How to cope with different degrees of cultural heuristics?**	**Which layers of cultural influence are addressed?**	**How to prevent a bias by the developer's cultural background?**	**How to get reliable data?**	**How to cope with the multiplicity of influences on behavior?**	**How to assess the results of integrating cultural aspects?**
Blanchard & Mizoguchi, 2008	By its constituents: norms, references, beliefs, behaviors, rituals, traditions, symbols, artifacts	Upper ontology for developing methodologies for assessing individual cultural profiles	Acknowledged, no solution	Identification of layers of influences like emotions, motivation or pedagogical strategies	Cultural bias is used as feature for authoring content	Upper ontology provides guidelines for eliciting data		
Nazir, Lim, Aylett, Cawsey & Enz, 2009	Hofstede (2001)	Synthetic cultures (Hofstede, Pedersen & Hofstede, 2002)		Action selection (e.g. rituals) and emotion appraisal		Literature on synthetic cultures	Action selection is influenced by emotional state, personality and cultural profiles	
Aylett, Paiva, Vannini, Enz, André & Hall, 2009	Hofstede (2001), Bennett's (1986) model of intercultural development	Defined by position on Hofstede's (2001) dimensions		Rituals and symbols		No reliable data necessary, fantasy cultures	Combination of emotional and cultural influence	Questionnaires
Miller, Wu, Vakili, Ott & Smith, 2009	Brown & Levinson (1987)	Language group		Perceptions, beliefs, attitudes, goals				Not clear how evaluation scenario is able to measure cultural aspects
Cassell, 2009	Cultural profile is created by common behavior patterns	Individual cultural profile defining cultural identity	Cultural background influences behavior based on contextual factors	Appearance, dialect, gaze		Observational study and analysis	Identification of factors that contribute to ascribing a specific cultural background	Principled variation of cultural features and rating by participants

Table 2. Continued summary of how the systems reviewed in this section address the identified pitfalls.

	Definition of culture	**How to cope with the multiplicity of the concept of culture?**	**How to cope with different degrees of cultural heuristics?**	**Which layers of cultural influence are addressed?**	**How to prevent a bias by the developer's cultural background?**	**How to get reliable data?**	**How to cope with the multiplicity of influences on behavior?**	**How to assess the results of integrating cultural aspects?**
Ruttkay, 2009	Culture as set of characteristics for group of people	National cultures are identified as problematic, other influences are apparent	Acknowledged, no solution	Emotional facial expressions	Acknowledged for labeling of data, no general solution	Literature and controlled experiments		
Koda, Ishida, Rehm & André, 2009		National cultures		Emotional facial expressions	Designer's cultural background is used as experimental condition	By exploiting the implicit knowledge of the designer		Perception studies based on implicit cultural knowledge of participants
Johnson and Valente, 2008	Culture is used as synonym for nation or language group			Appearance, language and pragmatics of its use, communicative strategies			Complex agent architecture with emotional appraisal, culture influences interpretation of user input	Evaluation of learning success, no information about transfer to real mission situations
Lane & Hays, 2008				Interpretation of the meaning of communicative acts		Literature		Evaluation of effectiveness of pedagogical strategies
Warren, Diller, Leung, Ferguson & Sutton, 2005		Unclear, seems to be national culture, but also subcultures like gamers/ non-gamers are mentioned		Decision making, coordination, task performance	Acknowledged but not addressed			

Table 3. Continued summary of how the systems reviewed in this section address the identified pitfalls.

	Definition of culture	How to cope with the multiplicity of the concept of culture?	How to cope with different degrees of cultural heuristics?	Which layers of cultural influence are addressed?	How to prevent a bias by the developer's cultural background?	How to get reliable data?	How to cope with the multiplicity of influences on behavior?	How to assess the results of integrating cultural aspects?
Khaled, Biddle, Noble, Barr & Fischer, 2006	Hofstede's (2001) dimension of identity		Not relevant because focus on typical behavior of group	Persuasion strategies	Utilized as feature for participatory design method	Literature, feedback from designer and user		
Isbister, Nakanishi, Ishida & Nass, 2000	Hall (1966)	National cultures		Appearance, communication topic (safe/unsafe)		Questionnaires		Questionnaires
Jan, Herrera, Martinovski, Novick & Traum, 2007	Something learned and controlled on subconscious level	Mix of different granularities, US vs. Mexican vs. mixture of Arabian cultures		Proxemics, gaze, overlap in turn taking		Literature	Acknowledged but not addressed	Perception experiments, rating of realism compared to own culture
Rehm, Nakano, André & Nishida, 2008	Hofstede (2001), Hall (1966)	National cultures	Probabilistic model	Appearance, verbal, non-verbal behavior, communication management	Observational studies in specific contexts, discussion of developers from target cultures	Observational studies, literature		Perception studies on isolated behavioral dimensions

interrelations between them that are necessary to describe and define culture-specific phenomena. In their endeavor they identify most of the above mentioned issues and present some ideas on how to deal with the arising challenges. They note that the notion of culture is often carelessly used in everyday conversation but hard to pinpoint down by a proper definition. By the means of their ontology, they define culture by its constituents in the form of norms, references, beliefs, behaviors, rituals, traditions, symbols, and artifacts. Thus, culture is a group phenomenon meaning that the behavior of individuals can deviate from these patterns. The cultural profile of an individual is multifaceted and depends on the individual's personal history. Their hope is that an upper ontology might be helpful in developing methodologies to assess such profiles. A number of different layers of cultural influences are identified like emotions, motivation or pedagogical strategies among others. An upper ontology provides structures for these different layers that can then be used to develop culture-specific ontologies for instance of Japanese pedagogical strategies vs. German pedagogical strategies (e.g. Hayashi, Bourdeau & Mizoguchi, 2008). The cultural bias of the developer on eliciting, interpreting, and structuring data is acknowledged and turned into a feature by using this cultural bias as a prerequisite in collaborative authoring and analysis of cultural data. Thus, different interpretations of the data can be provided and integrated. This allows ensuring the reliability of data.

Nazir and colleagues address the challenge that social behavior is influenced by multiple factors and present a first concept for a computational model that integrates emotions, personality and culture as influencing factors (Nazir, Lim, Aylett, Cawsey & Enz, 2009). Hofstede is taken as the general theory, but focusing on his idea of synthetic cultures used in training simulations (Hofstede, Pedersen & Hofstede, 2002). Thus, they are not dealing with actual cultures but parameterize influences on behavior by theoretical abstractions of the endpoints of Hofstede's dimensions. In their approach, culture influences action selection in the form of rituals (e.g. greeting) and emotional appraisal. Empirical data as such is not necessary in this model as only ideal (stereotypic) behavior is regarded that is linked to the extremes of Hofstede's dimensions. Their main contribution lies in the combined influence of the emotional state of the agent, its personality and its cultural profile on the action selection process.

Aylett and colleagues show how this kind of model can be employed to realize a system for increasing intercultural sensitivity in collaborative role-plays with embodied agents (Aylett, Paiva, Vannini, Enz, André & Hall, 2009). In contrast to the other systems presented here, their approach focuses on interactions within a non-realistic culture in order to prevent previously established stereotypes to interfere with the learning goals. A group of user's has to solve a task on a different planet and has to convince the inhabitants to cooperate with them. The underlying architecture of the alien agents extends a previously introduced architecture for empathic agents with a cultural level that relies on Hofstede's dimensions to represent different cultural backgrounds. The cultural background then influences the interpretation of behaviors and the action selection process. In a close analogy to the Chomskian ideas of language use, a universal behavior selection process is realized, which is augmented with culture-specific transformation rules for perceptions and actions. Cultural influences become apparent on two levels, rituals (i.e. scripts to be followed) and symbols. By symbols, emblematic gestures or cultural norms of behavior are meant (like bowing vs. handshake for greeting someone). As they are dealing with a fantasy culture, no reliable data is needed. If at all, data about the target audience is needed to prevent the agents from behaving to similar. The system has not been evaluated so far, thus it remains to be shown if this kind of role-play with fantasy cultures does lead to transfer effects to real intercultural encounters.

Miller and colleagues present a quite different approach to the challenge of modeling culture-specific behavior in interactive systems (Miller, Wu, Vakili, Ott & Smith, 2009). They consider Brown & Levinson's (1987) theory on politeness as the bridge between abstract cultural parameters like they are described by Hofstede and specific behavioral differences found in cross cultural comparisons. Using Brown & Levinson's theory as a foundation they develop a quantitative computational model of decision making that is influenced by culture-specific ways of expressing and interpreting politeness and thus reacting according to the given social context. Their argument is that theories like Hofstede's introduce culture as high level concepts, which creates the problem of relating to actual observable behavior. Thus, they propose politeness as a fundamental concept that mediates between universal cognitive processes and concrete realizations in specific contexts. Politeness theory is centered on the notion of face or – to be more precise – on keeping one's face. People maintain positive and negative face, which are continuously threatened during interactions, e.g., by commands or criticism on one's behavior. Brown and Levinson distinguish four different types of strategies to deliver such a threat. It can be delivered directly without any redress, which is the most rude form but sometimes inevitable, e.g., if it is crucial for the well-being of the addressee. Most of the time, speakers try to redress or mitigate such undesirable acts, e.g., by referring to the good looks of the addressee before asking her for a favor. Other strategies focus on the addressee's freedom of action and come e.g. in the disguise of apologies or impersonalizations. Brown & Levinson show that these strategies are universal but that every language has developed their own way of realizing these strategies. By relying on politeness theory, Miller and colleagues thus equate culture with language groups and accordingly focus on linguistic and cognitive behavior concerned with perceptions, beliefs, attitudes, and goals, which all influence the application of a given politeness strategy. They note that building the necessary databases for culture-specific behavior variations is difficult and fundamental for the task but present no solution for this problem. An evaluation is presented that seems not adequate to test for cultural differences in polite behavior. The participant is in the role of a dispatcher at a fire department and has to answer to requests. Social variables like status are manipulated for the requests, and the politeness of the participant's answer is measured. It is not clear how this scenario is able to measure cultural differences. Especially as the scenario implies urgency and thus the polite behavior might be distorted due to the context of the test setting.

Cassell introduces a yet another perspective on culture into the discussion (Cassell, 2009; Iacobelli & Cassell, 2008). Her approach is not based on a theory that defines culture by difference but rather by social practices resulting in no clear-cut groups like the American nation or the Japanese nation but emphasizing the fact that cultural identity is a multifaceted construct resulting from one's personal history of social encounters and changeable due to context like switching from the cultural habits of a board meeting to the cultural habits of a family dinner. Thus, the focus of their work is on examining how cultural identity is attributed to an agent. In a first study, they distinguish between the surface features of an agent (appearance) and its verbal and non-verbal behavior and present an experimental design to examine the relative importance of these factors to show which features are crucial for the attribution of a cultural background to an agent. In their series of experiments they prepared conversational agents with the outward appearance (clothing, skin color) of two US American subcultures, middle class white Americans and African Americans. Verbal behavior could be changed between Standard American English and African American Vernacular English. Different types of eye gaze were modeled as non-verbal behavior. They could show that the attribution of

culture to an agent can be triggered by its verbal and non-verbal behavior despite an appearance that suggests otherwise. Culture is defined in terms of ethnic identity which is created by common behavior patterns. Thus, there is not a single, monolithic cultural background of a person but it is necessary to observe "actual communities of people in particular contexts". The question if the cultural background of the developer influences the process is not addressed but the question arises if the choices of the parameters that are analyzed and interpreted in the observational studies are not already culturally dependent. The main contribution lies in the principled approach to identify the factors that contribute to attributing a specific cultural background to an agent. This is examined with varying the cultural features under investigation and rating the results by participants of the study.

Different work concentrates on culturally adequate facial expressions. Ruttkay (2009) focuses on emotional facial expressions and analyzes how culture-specific displays of facial expressions can be created and how the cultural background of the user may influence the perception of such expressions. Culture is defined as "a set of characteristics which form a 'common denominator' among groups of people". Cultural influences are thus apparent on different levels ranging from values in life to multimodal behavior. The use of national cultures as the basic concept is shown as too stereotypical as a system will not interact with the average Japanese but with a fisherman from Hokkaido or a professor of Informatics from Tokyo. Consequently, Hofstede's approach of national culture is identified as problematic because other influences are apparent like subcultures, age, gender, and many more. From the multitude of cultural influences, she focuses on emotional facial expressions and acknowledges the difficulty of reliably identifying the relevant features, describing cultural influences and labeling observational data due to the implicit bias of the individual cultural profiles of both the observed subjects and the expression coders. No general solution is offered for this dilemma. In order to acquire reliable data on which to ground the ECA modeling, a mix of exploiting the literature and controlled experiments is suggested. For the area of facial expressions, a mapping between facial displays and emotions is identified and can be used to design principled elicitation experiments.

Koda and colleagues present a series of experiments that investigate culturally determined interpretation of emotional facial expressions by avatars (Koda, Ishida, Rehm & André, 2009; Koda, Rehm & André, 2008). To this end, professional designers from Japan, the US, France and the UK create twelve static expressions of facial expressions for avatars in the predominant comic style of their culture. Participants in the experiments had to assign twelve emotion labels to these expressions. The results show that perception differs across cultures, and that an in-group advantage exists for correctly interpreting the expressions, which means that e.g. facial expressions created by Japanese designers are more easily recognizable by Japanese participants than by e.g. French. Culture is interpreted as national culture without resorting to a more elaborate theoretical background. The focus lies on (artistic) emotional facial expressions. The inherent bias of the designer's cultural profile is used as a controlled experimental condition. The designer creates the avatar expressions making use of his cultural heuristics. The perception then is guided by the participant's cultural background, which again is utilized as a condition in the experimental design. The design follows a paradigm by Elfenbein and Ambady (2003) in distinguishing between expressers (culturally determined avatar design) and recognizers (participants in the experiment), which allows to exploit the implicit cultural knowledge of all participants.

The tactical language training system is the most advanced and commercially successful ITS which takes cultural issues into account (see Johnson & Valente (2008) for a recent overview). It provides the learner with role-playing experi-

ences in games with virtual characters that are tailored to the context of specific cultures (e.g. Iraqi, Afghan) and military scenarios. Main goal is the acquisition of language skills and some corresponding cultural behaviors like appropriate use of honorifics or knowledge about gestural emblems. Interaction is speech driven accompanied by menu-based selections, for instance for appropriate gestures. In this approach, the notion of culture is used as a synonym for a language group aka nation without resorting to an elaborate theoretical approach of defining culture. Cultural influences that are addressed implicitly and explicitly are the appearance (of agents), language (phrases, honorifics, etc.) and the pragmatics of its use as well as communicative strategies. To realize the role-playing game, a complex agent architecture has been developed that integrates emotional appraisals of situations. Cultural influences are modeled as an additional layer that acts as a pattern recognition mechanism, interpreting the input of the user in terms of right or wrong usage of cultural behavior patterns. The learning success regarding language learning has been evaluated but no information is so far available about the transfer of the assumingly acquired cultural knowledge to real-life mission situations.

Lane and Hays (2008) describe an intelligent tutoring system tailored to coaching business etiquette in intercultural encounters. Similar to the tactical language training system, users are provided with a role-playing game and a negotiation mission and have to adhere to cultural rules of social conduct in order to achieve their goals, e.g. to not offer alcohol as a present in an Arabian setting. The main focus of their approach is on the pedagogical effectiveness of the ITS, leaving most of the questions raised here unanswered as culture is more like the back story of the interactive narrative. Thus, the definition of culture is not directly addressed but equated with nations. Cultural influence is located on an abstract cognitive level relating to the interpretation of the meaning of communicative acts and the "willingness to assume a different cultural perspective". Additionally, appearance and nonverbal behavior like gestural activity are identified as other layers of cultural differences but not directly addressed. The culturally different rules and behaviors are treated as data for the system, thus providing the relevant data is a problem of knowledge representation. The data for the current version of the system was gained from a literature review. As culture is merely the back story for the ITS, the evaluation focuses on the effectiveness of pedagogical strategies and not on how realistic the modeling of cultural differences in negotiation behavior is.

That the neglect of the issues raised here can result in quite dubious systems is exemplified by work of Warren and colleagues (Warren, Diller, Leung, Ferguson & Sutto, 2005). They propose using a commercial role-playing game as an environment for research on social and cultural aspects of communication focusing especially on collaborative team work. They strongly argue for the feasibility of this approach and its primacy over simple pencil and paper studies due to the immersive and task-oriented nature of simulation games. Unfortunately, the claim for usefulness especially in relation to cultural aspects of communication is not corroborated beyond some general remarks. Especially doubtful is their pilot study on cultural effects of negotiation. The approach is problematic because i.) the game itself is culturally biased as it is a typical Western military action game, ii.) the creation of the two groups that are compared is invalid as they compare US American teams with multinational teams, and iii.) possible decisions in the game seem to be solely based on the developers intuition and thus their cultural background. They claim to investigate the cultural impact on decision making, coordination, and performance. But they fail to identify what they mean by culture and how it influences the identified processes. Their work is not based on a cultural theory but seems to be working with national cultures. But then again they even mix those in their pilot study making it impossible to

attribute the results of the study to cultural effects. Later they present two subcultures (gamers vs. non gamers) and analyze their impact on behavior without noticing it.

Khaled and colleagues present a persuasive game that is specifically tailored to collectivist cultures (Khaled, Biddle, Noble, Barr & Fischer, 2006). In their analysis of current persuasive technology they discovered a lack of culturally tailored persuasive strategies. All systems seem to target an audience from individualistic countries focusing on individual motivations and conditioning. Based on the literature on differences between collectivistic and individualistic societies, Khaled and colleagues defined persuasion strategies that take the group focus into account and present a smoking cessation game which incorporates these strategies and is targeted to the (collectivist) Maori culture. In their approach, they regard individualism vs. collectivism as the crucial dimension of culture following Hofstede for defining this dimension. To realize their game, they rely on the definition of typical behavior of a group, not focusing on individual variance in this behavior. They focus on the definition of persuasion strategies for collectivist cultures, acknowledging the problem of the cultural bias of developer and user and converting this bias into a feature of the design process. To this end, they suggest a participatory design method for developing the content of the game. Thus, the necessary data is acquired by exploiting the information from the literature and integrating the feedback from the developers and users.

Isbister and colleagues present an agent that is designed to further discussions between participants from different cultures and help establishing a positive impression of the communication partner (Isbister, Nakanishi, Ishida & Nass, 2000). To this end, they created an agent that overhears video chat conversations between US and Japanese students and intervenes when it detects longer silences in the conversation. In this case the agent tries to elicit a new (safe) topic for continuing the interaction and withdraws afterwards. Although the system tries to further cross-cultural communication it also exemplifies many of the pitfalls in enculturating conversational agents. Again, culture is not defined theoretically but by nationality. Hall's theoretical approach (e.g. Hall, 1959) is referenced but only to underline that differences in social behavior exist between US Americans and Japanese. Cultural differences in communication behavior are assumed, implicitly concerning the appearance and perception of the agent (comic-style dog) and explicitly concerning safe/unsafe topics in conversation. The agent has been realized as a comic-style dog in order to induce an image of friendliness and helpfulness with a drawing style somewhere between American comics and Japanese mangas. The cultural bias of the developer is not addressed in this approach and might be one reason for the performance of the system. It worked for US students as expected and showed nearly opposite results for the Japanese students. Features of the system that might be implicitly tailored to the Western perspective include, apart from the language (English), the interpretation of pauses as awkward, and the order of topics. It has been shown (e.g. Endrass, Rehm & André, 2009) that pauses are perceived as less problematic in the Japanese culture. Thus, the definition of pauses as the only feature to intervene could have been interpreted by the Japanese participants as unwanted interruptions. Additionally, it has been shown that the order in which topics are discussed differs between individualistic (US) and collectivistic (Japan) cultures (e.g. Hall, 1966). As this is not an explicit feature of the system, it can be guessed that the agent adheres to a Western order, which again might be a cultural bias in favor of the US participants in the study. To gain reliable information on safe topics in conversation, questionnaires have been used in both cultures and only those topics that count as safe in both cultures have been integrated in the system.

Jan and colleagues present a parameter based model to modify certain behaviors of virtual agents in a culture-specific way (Jan, Herrera, Martinovski, Novick & Traum, 2007). To this end, they extend a personality-based computational model in order to realize constraints for the agent animations. A behavior profile for a given culture is specified in terms of proxemics, gaze and turn-taking using an XML structure. They define culture as something that is controlled and learned on a subconscious level but refrain from resorting to a precise definition of culture. This results in the problem of a different granularity of the cultural aspects that are modeled, bearing some similarity with the work by Warren and colleagues. Jan and colleagues introduce the US and the Mexican culture and an Arabian culture that incorporates a number of different nations (Lebanon, Qatar, Syria, Kuwait, Palestine, Morocco, Egypt) with quite different historic, geographical and political backgrounds. Cultural influences are taken into account based on the ease of integration in an existing agent framework and include proxemics, gaze and overlap in turn taking (although this is not tested during the evaluation as the evaluation works with silent movies). The necessary data is drawn from a literature review. It turns out that the information from the literature is in most cases merely qualitative in nature, often gives only mean values (for instance on proxemics) or does not give information about a culture under investigation (for instance turn-taking in Arabian cultures). Coping with this lack of data can lead to inconsistencies. For instance, while differences in the means of proxemics behavior were used for extrapolating all spatial behavior for the Arabian and Mexican cultures, no data at all could be obtained for Arabian turn-taking behavior. Thus, turn-taking is not modeled specifically for this culture. A consequence of this procedure is a mix of culture-specific behavior, in this case American turn-taking/overlap with Arabian proxemics and gaze. For evaluating the effects of culture-specific behaviors, Jan and colleagues present a series of perception studies that rely on movies with groups of characters differing in proxemics and gaze behavior and pauses in turn-taking. Subjects are asked to rate the realism of the animations with respect to their cultural background. Additional questions concerned elements that they thought as inappropriate in their culture. Results indicate differences in the perception of differences, i.e. subjects with Arabian backgrounds identified differences in behavior in a different way than American and Mexican subjects.

The CUBE-G project by Rehm and colleagues combines a bottom-up data-driven approach with a top-down model-driven one in order to develop an agent architecture that allows parameterizing the cultural background of the agent in order to adapt its behavior to the user's cultural background (e.g. Rehm, Nakano, André & Nishida, 2008). To this end, they collected a large corpus of multimodal behavior in situations that every ex-patriate or even tourist may encounter (first meeting, negotiation, status difference). By analyzing the non-verbal behavior in these situations they are able to define a probabilistic model of behavior that can be used to interpret the user's input and to render the behavior of an agent. Hofstede's theory presents the basic theoretical construct accompanied by Hall's (1966) model of proxemics. In order to allow variations in the culture-specific behavior, a probabilistic model is introduced. Rehm and colleagues focus on a broad range of cultural influences like appearance, non-verbal behavior (proxemics, sound, gestural activity, posture), verbal strategies, and communication management (turn-taking). To keep the cultural background of the developers at check, a close cooperation between researchers from the target cultures is maintained. Data is gained by standardized observational studies that are described in more detail below. An evaluation strategy is proposed that relies on perception studies that vary isolated behavior traits which are then rated by participants from different cultures.

Tables 1, 2 and 3 summarize the features of the systems discussed here. What is apparent is the lack of a cultural theory or model in a number of systems. This is a crucial problem, as these systems rely on unstated assumptions on what constitutes a culture and its influences on (system) behavior. But without a proper theoretical background the integration of cultural aspects of behavior and its interpretation in a system can be only descriptive and superficial at most. For an engineering approach like the development of an ECA is, a theoretical background is necessary that clearly states (i) how to assess cultural values and norms or how to assess an individual's cultural profile and (ii) the relation between cultural features and behavioral heuristics. Three trends have been identified above: (i) theories that define culture by norms and values, thus highlighting the differences between groups, generally in the form of nations and thus establishing culture as a social group phenomenon, (ii) "linguistic" models that see language as the crucial determinant of culture and thus equate cultural groups with language groups, and (iii) theories of cultural practice that highlight the fact that people are embedded in a multiplicity of groups with context-dependent shared common values and thus culture is an individual trait, a person's cultural identity is dependent on personal history and social environment. The majority of current systems is based on Hofstede's dimensional model, Hall's dichotomies or equates culture and language. The obvious appeal of a dimensional model like Hofstede's is that it allows defining national cultures in the form of attribute value pairs that then can be linked to information found in the literature about behavior differences between national cultures like Japanese vs. US American vs. Italian. Thus, user groups can be categorized according to their national affiliation and it suffices to develop systems that are able to adapt to a "reasonable" number of different conditions. The equation of culture with language groups stems from the attempt of developing tutoring systems for language learning and supplies evidence for the use of general cultural models like Hofstede's. In the case of language learning, one is primarily not interested in teaching for instance a variant of German that is spoken by Bavarian farmers but a standardized German which is as fictional as the idea of a standardized German culture. Nevertheless, nobody would argue that this way of teaching a language has not its merits and works quite well. Thus, it remains to be shown that a system that equally idealizes the cultural heuristics of a national or a language group does not have its merits for enculturating interactive systems. Although the specific behavior of individuals is context- and situation-dependent and influenced by their personal history, the idea is that stable traits can be distinguished that are attributable to a given cultural group. Such traits are not necessarily apparent in a given situation but would be perceived as "normal" behavior by members of that group. Consequently, a definition of culture as a group phenomenon is thus a good starting point to the endeavor of enculturating ECA systems.

On the practical level, a plethora of cultural influences has been integrated, ranging from appearance of the agents (skin color, dress) over verbal (language, turn-taking behavior, dialogue strategies, etc.) and nonverbal behavior (gaze, gestures, proxemics, etc.) to cognitive processes (appraisal of input, decision making, etc.). This exemplifies the complexity of ECAsystems that allow for complex multimodal input and output behavior and at the same time raises the question how the integration of cultural aspects can reliably evaluated with the resulting system as it is apparent that not all cultural influences can be taken care of at once. Perception studies with members of the target culture are presented as one method for this evaluation but it remains unclear how the effects of cultural modeling can be distinguished from general effects of the ECA system.

THE DEVELOPMENT PROCESS OF ECA SYSTEMS

The methodological approach for modeling the behavior of Embodied Conversational Agents is well exemplified by the following development steps:

- Study: To build a formal model for generating realistic agent behaviors, data of human interactions is necessary for two reasons: (i) it serves as an empirical foundation for the formal models of human agent interaction, and (ii) it serves as a benchmark against which these models are evaluated. In most cases, formal models are not built from scratch. Rather, the data analysis serves to refine existing models found in the literature. Such models often lack explicit information necessary for the integration in an agent system like synchronization and timing of modalities. Over the last decade, numerous work has established the area of multimodal corpus analysis to shed light on the specifics of multimodal interaction. To give some examples, Kipp, Neff, Kipp & Albrecht (2007) suggest an annotation scheme for gestures that draws on the distinction between the temporal course of a gesture and its type and relies on a gesture typology introduced by McNeill (1992). Chafai, Pelachaud & Pelè (2006) annotate instead the expressive dimensions of gestural activity focusing on how a gesture is accomplished and not on what kind of gesture is used. Rehm & André (2007) describe an annotation scheme that analyzes gestures on a more abstract functional level. Their corpus captures the relation between linguistic and nonverbal strategies of *politeness.*
- Model: The data gathered in the previous step of the development process serves as the foundation of a formal model of human agent interaction. Caridakis and colleagues give an account on how the data from such a corpus can be used to directly mirror the behavior of a human speaker with an agent (Caridakis, Raouzaiou, Bevacqua, Mancini, Karpouzis, Malatesta & Pelachaud, 2007). A similar approach is described by Kipp, Neff, Kipp & Albrecht (2007), who extract information of personal idiosyncrasies of the human speaker, which is then mimicked by the agent. Lee & Marsella (2006) extract statistical rules from a corpus of natural dialogues that allow them to generate appropriate head and hand gestures for their agent that accompany the agent's utterances. Instead of rules, Rehm & André (2007) have shown how statistical information can be extracted from a multimodal corpus and used as control parameters for a virtual character. To this end, they analyzed what kind of relation exists between certain types of gestures and verbal strategies of politeness. The resulting models of human-human conversational behavior then serve as a basis for the implementation of ECAs that replicate the behaviors addressed by the models.
- Test: To evaluate the resulting system, experiments are set up in which humans are confronted with ECAs following the model. The data collected in the first step can serve as a baseline against which the resulting ECA implementation can be tested. Nakano and colleagues exemplify this use of multimodal corpora in developing agents that exhibit human turn taking and grounding behavior (Nakano, Reinstein, Stocky & Cassell, 2003). Rehm (2008) uses instead a corpus of human agent interactions to exemplify how design guidelines can be derived on this basis for such interactive systems.

The above mentioned work concentrates on the challenge of realizing natural interaction behaviors for agent systems but did not acknowledge culture as a relevant parameter that might influence such interactions. Based on this general development process, the next section introduces strategies for introducing culture into the development process and draws from examples of the CUBE-G project that aims at developing a parameterized model of cultural influences on the verbal and non-verbal behavior of Embodied Conversational Agents.

Culture in the Development Process of ECA Systems

In the previous sections, we have introduced the challenges of integrating culture into the development process and ultimately into the ECA system. In this section, strategies are introduced how to avoid the pitfalls of this multi-defined domain in relation to the general development process of ECA system.

Study

Above we have argued for Hofstede's theory of cultural dimensions as a starting point for enculturatingECA systems. Although the theory describes certain correlations between cultural dimensions and correlated behavioral heuristics, this attribution is not unambiguous as the correlated heuristics might contradict each other on different dimensions. Consider for instance the following example dealing with proxemics. High power distance (hierarchy dimension) might result in standing further apart in face-to-face encounters whereas collectivism (identity dimension) generally means standing closer together in the same situation. Both attributions hold true for the Japanese culture. Thus, what will be the result of these correlations if they are combined? Solutions of different complexity can be thought of. Interlocutors could position themselves simply in a mean distance. Or we could define a hierarchical relation between the dimensions resulting in some information being overridden or weighted differently. More sensible would be a contextual adaptation that takes the semantics of the dimensional position into account. If a culture has a high power distance then there could be differences in proxemics behavior that are related to social status, for instance resulting in standing further away from high status individuals but closer together with peers.

What is apparent from this example is one obvious conclusion. To adapt the behavior of agents to cultural heuristics it is indispensable to gain insights into how these differences manifest in face-to-face encounters. Unfortunately, there is a lack of reliable cross-cultural data as the information in the literature is often of an anecdotal character, or lacks technical information that is necessary to realize an interactive system. One way to deal with this problem is to gather data in a standardized way, tailored to the modeling endeavor. Rehm and colleagues analyze in detail how this standardization process has to be organized for recordings across cultures (Rehm, André, Bee, Endrass, Wissner, Nakano, Lipi, Nishida & Huang, 2009). In order to define relevant scenarios, literature on cross-cultural training can serve as a guideline (e.g. Landis, Bennett & Bennett, 2004). Although qualitative in nature, it defines the key dimensions on which behaviors vary and focuses on those factors that tend to result in irritations between interaction partners from different cultures. Based on this type of information, Rehm and colleagues define three specific scenarios that provide the recording context and are likely to be relevant in every culture: (i) a first meeting, (ii) a negotiation, and (iii) an interaction with a person with higher social status. To ensure the replication of conditions in all cultures participating in the study, a common protocol was established on how to conduct the study with detailed instructions to be followed at every step. These instructions had to cover recruiting of subjects and actors, the timeline of each recording as well as "scripts" for

the people conducting the experiment as well as detailed information about the necessary materials and the setup of the equipment. To produce comparable data sets it was indispensable to define technical requirements for the video recording sessions. This included the specifications for the recording equipment as well as the layout of the recording area to be able to reproduce the recording conditions. To control for gender effects, a male and a female actor were employed in each scenario interacting with the same number of male and female subjects. For each subject, around 25 minutes of video material was collected, 5 minutes for the first meeting, 15 minutes for the negotiation, and 5 minutes for the status difference. To be able to control for effects of personality on the behavior under examination, participants had to fill out a NEO-FFI personality questionnaire (McCrae & Allik, 2002).

Whereas sometimes the developer's intuition might work due to the fact that the developer can take his own actions as a model for building the interactive behavior of an ECA, this is quite problematic if designing for a different culture. The developer's own cultural norms and heuristics hinder this process in making quite specific aspects of behavior relevant that might be irrelevant in a different culture. Consider the following example. If studying turn-taking behavior in Germany, the effect will be to consider an ordered exchange between interlocutors with little overlap as the basic form of discussion. But in other cultures it is more common to have strong overlaps and simultaneous turns in discussions to emphasize one's interest in the topic (Ting-Toomey, 1999). Thus, investigating turn-taking behavior in Italy might result in a completely different model of turn-taking behavior. Even when being aware of cultural differences, this does not necessarily help in identifying relevant behaviors. An obvious solution to this problem would be to always involve developers from the target cultures in the development process. This might only be feasible for large-scale projects. A low-budget solution would be to discuss most of the design choices as often as possible with someone from the target country. To do so, it is important to make one's own design choices explicit. As the underlying heuristics are implicit and generally interpreted as the "natural" way to do things, this might not be easy. To solve this problem, it seems inevitable to develop best-practice advices or guidelines on how to check for cultural issues in the design of interactive systems.

In the case of the CUBE-G corpus, the video data has been analyzed following standard annotation schemes that both the German and the Japanese project partners agreed on. Additionally, it is necessary to ensure that the coding of phenomena is reliable and comparable across different coders, which in our case come from different cultures. To ensure a consistent and reliable coding process, the coding schemes are evaluated regarding their ease-of-use and their quality (Dybkjaer & Bernsen, 2004). Ease-of-use describes the practical applicability of the scheme and includes interviews or questionnaires of the coders, verifying consistent use of tags by different coders, and measuring the time necessary for coding. Whereas the ease-of use criterion measures how coders can handle the scheme practically, the quality criterion measures if a coding scheme is suitable at all for its intended end. Quality is measured by coverage, reliability and consistency. Comparing different corpus samples allows to assess the coverage of the scheme i.e. helps us to determine whether it does code the features it was intended for. Inter-coder reliability is usually assessed by means of the kappa value, which takes into account deviations between the coder as well as an agreement by chance. A high kappa indicates that coders do not need much subjective interpretation of the phenomena. If a master code is available against which the results of coders can be matched, precision and recall are two additional measures. Precision just measures the proportion of correctly coded items whereas recall measures the proportion of coded items. Consistency at last

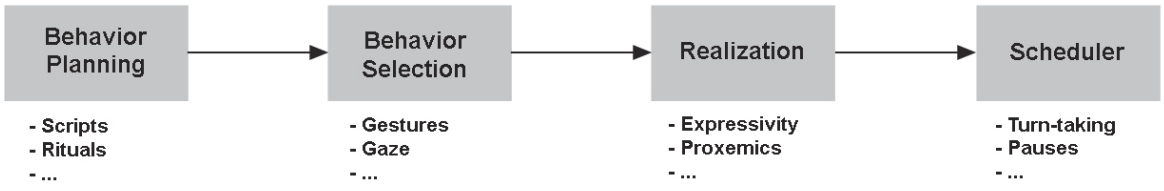
is assessed by comparing the results of the same coder for the same sample at different times. Apart from these general quality assessments, analyzing data across cultures introduces another level of complexity. Defining a ground truth for such an annotation of cultural dependent factors of behavior is quite difficult. Thus, letting German coders evaluate annotations of the Japanese interactions and vice versa gives additional information on the perception of prototypical situations and behavior in the two different cultures that would otherwise go unnoticed and provides another, cultural, level of evaluating the coding schemes.

Model

If we roughly sketch the process of behavior selection and generation in an agent system, it becomes obvious that culture penetrates most stages of this process. Figure 2 gives a simplified impression of some of the main processing steps. In the planning stage, culture provides scripts and rituals for interactions. One of the most fundamental situations in this respect is a first meeting encounter which serves as an example here. According to Argyle (1975), a first meeting is a ritual that follows predefined scripts. Ting-Toomey (1999) follows this analysis by denoting a first meeting as a ceremony with a specific chain of actions. Behavior selection is concerned with enriching the dialogue step with suitable verbal and nonverbal behavior. Consider the use of gestures as an example. Culture influences the selection process on different levels. On the one hand, it is necessary to choose the right gesture type and animation for the utterance. This repertoire of available gestures is at least partially culture-specific as there are sets of language and thus culture-specific emblematic gestures. On the other hand, if and how many gestures are employed in an utterance differs widely between cultures. The Italian culture for instance has a rich repertoire of emblematic gestures and gestures in general are used frequently in face-to-face encounters. Quite the opposite is true for the German culture. In the realization stage another influence of culture comes into play. Consider again gestural activity. Whereas one culture gestures fast and frequently, taking much space in doing so, other cultures make only use of infrequent gestures that do not intrude the space of the interlocutor (Efron, 1972). The scheduling stage at last is necessary to ensure appropriate timing in turn-taking of the interlocutors, which again is culture-specific. For instance in the above mentioned study on German and Japanese behavior, we found that German interlocutors are generally uncomfortable with longer pauses in conversations compared to the Japanese samples (Nakano & Rehm, 2009).

One suggestion to deal with this ubiquitous influence of culture and especially with the fact that cultural heuristics are only "guidelines" to behavior but that the actual behavior of a given person in specific contexts can vary substantially is presented in Rehm, Bee & André (2008). By modeling the causal relations between a culture's location on Hofstede's dimensions and correlated behavior in a probabilistic Bayesian network, it becomes possible to categorize the user's behavior

Figure 2. Cultural influences during the behavior selection process.



in terms of cultural profiles, which in turn allows setting the agent's cultural profile and inferring behavioral heuristics for this profile. Causes in this model are then the positions of a culture on the single dimensions and corresponding effects are observable behaviors like speed or spatial extent of gestures. The user's observed behavior is set as evidence for the output nodes of the Bayesian network (Rehm, Bee, and André 2008). A diagnostic inference then yields the most likely causes, i.e. the most likely positions on Hofstede's dimensions, which again can be used to infer the user's cultural group. Additionally, making use of the cultural dimensions allows abstracting from the specific culture of the user to a distribution on the five dimensions. Thus, deviating behavior of the user, i.e. behavior that is not in accordance to known patterns of behavior for the user's culture, results in a different interpretation of the single user's position on the cultural dimensions, capturing the effect that cultural patterns of behavior are group phenomena and that individuals can deviate from these heuristics based on their individual cultural profiles. It remains to be shown if the user is then irritated by the system's behavior which is not in accordance with his "real" cultural background. When setting the agent's nonverbal behavior, the Bayesian network delivers information about dominant patterns of behavior in a culture that is found at the corresponding locations of the cultural dimensions, for instance low on hierarchy, low on identity, high on gender, medium on uncertainty, high on orientation. This results in a probability distribution for each behavior e.g. for volume of speech the probabilities could be 70% high, 29% medium, and 1% low. This probability distribution can now be utilized to set the agent's volume of speech in the next dialogue act.

Like in the previous development step, the cultural background of the developer supplies heuristics on what is interpreted as relevant or typical behavior. At this point, this check is necessary on different layers of abstraction. The developer's background may bias how the data derived in the previous step is used to model the behavior of an ECA. The definition of objective criteria is a necessary prerequisite for a reliable analysis. Actually building the ECA based on the analysis and the model suffers from the same pitfalls as before. What is an unimportant variation in gestural expressivity in one culture might lead to severe misunderstandings in a second culture. The same suggestions that were presented in the previous section apply here, i.e. intensive discussions with members of the target culture are mandatory. In the long run, the development of best-practice guidelines seems inevitable.

Test

Having enculturated the ECA system with cultural aspects it remains to be shown how successful this integration is. This is not easy because, as we have seen above, cultural heuristics serve as general guidelines but may not be applicable to the same degree in every situation. Moreover, culture is not the only influence on behavior, thus the user's reaction to the agent might not be attributable to a cultural effect (alone). The lack of principled evaluation methods for enculturated systems makes it difficult to compare the performance of systems that claim to take cultural aspects into account. In order to do so, evaluation methods have to be developed focusing on different aspects of the enculturation process of ECA systems:

1. To determine if the implemented model of cultural influences produces behavior that is relevant to establish a certain cultural identity of the agent
2. To evaluate the effects of integrating culture on the human agent interaction, (i.e. to show the fitness of the agent in terms of the modeling goals), which can be to:
    a. Increase cultural awareness on the side of the user
    b. Model behavior for educational purposes

   c. Adapt to the user's background for more efficient communication
3. To evaluate if the user's cultural background can reliably be predicted observing his interactive behaviors

The last item is necessary if the agent should be able to autonomously adapt to the cultural background of the user. Perception studies seem to present the best method so far to test for isolated effects (spatial extent of gesture vs. speed of gesture) as well as combinations of agent behaviors (spatial extent and speed of gestures). Ideally, such an approach allows for establishing a ranking of important factors. Above, the work of Iacobelli and Cassell (2008) presented an example that showed the relative importance of behavior over appearance of the agent. To evaluate if the integration of cultural aspects really has an effect on the task, performance with the enculturated agent has to be compared with a standard version or with performance in the human-human condition. To this end, the corpus that was recorded in the study step of the development process can serve as a benchmark for performance and can be utilized for the comparison.

CONCLUSION

The chapter introduced the pitfalls that await the developer of enculturated conversational agents and presented a number of strategies or best-practice advices how to cope with this multi-defined domain during the development process. These strategies and advices are not solely tailored to Embodied Conversational Agents but may be of a more general interest for the integration of cultural aspects in interactive systems. The review of current approaches, which embrace this challenge, has shown the great potential of this endeavor for instance in the area of intelligent tutoring systems or persuasive technologies. It has also shown the lack of a common theoretical background and research methodology. This line of research is still in its infancy and the strategies presented in this paper may help serving as guidelines for developing the necessary methodological approach to tackle the challenge of enculturating ECAs and HCI in general.

ACKNOWLEDGMENT

The work described in this paper was supported by the German Research Foundation (DFG) with research grant RE2619/2-1.

REFERENCES

Argyle, M. (1975). *Bodily Communication*. London: Methuen & Co. Ltd.

Aylett, R., Paiva, A., Vannini, N., Enz, S., André, E., & Hall. L. (2009) But that was in another country: agents and intercultural empathy. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*.

Bennett, M. J. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations*, *10*(2), 179–195. doi:10.1016/0147-1767(86)90005-2

Blanchard, E. G., & Mizoguchi, R. (2008). Designing Culturally-Aware Tutoring Systems: Towards an Upper Ontology of Culture. In Blanchard, E. G., & Allard, D. (Eds.), *Culturally-Aware Tutoring Systems (CATS2008)*. Montréal, Canada.

Brown, P., & Levinson, S. C. (1987). *Politeness – Some Universals in Language Usage*. Cambridge, UK: Cambridge University Press.

Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L., & Pelachaud, C. (2007). Virtual agent multimodal mimicry of humans. *Language Resources and Evaluation*, *41*, 367–388. doi:10.1007/s10579-007-9057-1

Cassell, J. (2009). Social Practice: Becoming Enculturated in Human-Computer Interaction. In Stephanidis, C. (Ed.), *Universal Access in HCI, Part I, HCII 2009* (pp. 303–313). Berlin: Springer.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents*. Cambridge, MA: MIT Press.

Chafai, N. E., Pelachaud, C., & Pelè, D. (2006). Analysis of gesture expressivity modulations from cartoon animations. In *Proceedings of the LREC Workshop on Multimodal Corpora*.

Damian, I., Janowski, K., & Sollfrank, D. (2009). Spectators, a joy to watch. In *Intelligent Virtual Agents (IVA)* (pp. 558–559). Berlin: Springer. doi:10.1007/978-3-642-04380-2_90

Dybkjaer, L., & Bernsen, N. O. (2004). Recommandations for natural interactivity and multimodal annotation schemes. In *Proceedings of the LREC'2004 Workshop on Multimodal Corpora,* Lisbon, Portugal (pp. 5-8).

Efron, D. (1972). *Gesture, Race and Culture*. New York: Mouton and Co.

Elfenbein, H. A., & Ambady, N. A. (2003). When Familiarity Breeds Accuracy: Cultural Exposure and Facial Emotion Recognition. *Journal of Personality and Social Psychology*, *85*(2), 276–290. doi:10.1037/0022-3514.85.2.276

Endrass, B., Rehm, M., & André, E. (2009). Culture-specific communication management for virtual agents. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 281-287).

Hall, E. T. (1959). *The Silent Language*. New York: Doubleday.

Hall, E. T. (1966). *The hidden dimension*. New York: Doubleday.

Hall, L., Woods, S., Aylett, R., Newall, L., & Paiva, A. (2005). Achieving Empathic Engagement Through Affective Interaction with Synthetic Characters. In Affective computing and intelligent interaction.

Hayashi, Y., Bourdeau, J., & Mizoguchi, R. (2008). Structurization of Learning/Instructional Design Knowledge for Theory-Aware Authoring Systems. In Woolf, B. P., Aïmeur, E., Nkambou, R., & Lajoie, S. (Eds.), *Intelligent Tutoring Systems* (pp. 573–582). Berlin: Springer. doi:10.1007/978-3-540-69132-7_60

Hofstede, G. (2001). *Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Thousand Oaks, CA: Sage Publications.

Hofstede, G. J., Pedersen, P. B., & Hofstede, G. (2002). *Exploring Culture: Exercises, Stories, and Synthetic Cultures*. Yarmouth, ME: Intercultural Press.

House, R. J., Hanges, P. J., Javidan, M., Dorman, P., & Gupta, V. (2004). *Culture, Leadership, and Organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage Publications.

Iacobelli, F., & Cassell, J. (2007). Ethnic identity and engagement in embodied conversational agents. In *Intelligent Virtual Agents (IVA)* (pp. 57–63). Berlin: Springer. doi:10.1007/978-3-540-74997-4_6

Isbister, K., Nakanishi, H., Ishida, T., & Nass, C. (2000). Helper Agent: Designing an Assistant for Human-Human Interaction in a Virtual Meeting Space. In *Proceedings of CHI* (pp. 57-64).

Jan, D., Herrera, D., Martinovski, B., Novick, D., & Traum, D. (2007). A Computational Model of Culture-Specific Conversational Behavior. In Pelachaud, C. (Eds.), *Intelligent Virtual Agents (IVA)* (pp. 45–56). Berlin: Springer. doi:10.1007/978-3-540-74997-4_5

Johnson, W. L., & Valente, A. (2008). Tactical Language and Culture Training Systems: Using Artificial Intelligence to Teach Foreign Languages and Cultures. In *Proceedings of IAAI 2008*.

Khaled, R., Biddle, R., Noble, J., Barr, P., & Fischer, R. (2006). Persuasive interaction for collectivist cultures. In W. Piekarski (Ed.), *The Seventh Australasian User Interface Conference (AUIC 2006)* (pp. 73-80).

Kipp, M., Neff, M., Kipp, K. H., & Albrecht, I. (2007). Towards Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In Pelachaud, (Eds.), *Intelligent Virtual Agents (IVA)* (pp. 15–28). Berlin: Springer. doi:10.1007/978-3-540-74997-4_2

Kluckhohn, F., & Strodtbeck, F. (1961). *Variations in value orientations*. New York: Row, Peterson.

Koda, T., Ishida, T., Rehm, M., & André, E. (2009). Avatar Culture: Cross-Cultural Evaluations of Avatar Facial Expressions. *AI & Society*. *Special Issue on Enculturating HCI*, *24*(3), 237–250.

Koda, T., Rehm, M., & André, E. (2008). Cross-cultural evaluations of avatar facial expressions designed by Western and Japanese designers. In Prendinger, H., Lester, J., & Ishizuka, M. (Eds.), *Intelligent Virtual Agents (IVA)* (pp. 245–252). Berlin: Springer. doi:10.1007/978-3-540-85483-8_25

Landis, D., Bennett, J. M., & Bennett, M. J. (2004). *Handbook of Intercultural Training*. Thousand Oaks, CA: Sage Publications.

Lane, H. C., & Hays, M. J. (2008). Getting down to business: Teaching cross-cultural social interaction skills in a serious game. In Workshop on Culturally Aware Tutoring Systems (CATS) (pp. 35-46).

Lee, J., & Marsella, S. (2006). Nonverbal Behavior Generator for Embodied Conversational Agents. In Gratch, J. (Eds.), *Intelligent Virtual Agents (IVA)* (pp. 243–255). Berlin: Springer. doi:10.1007/11821830_20

McCrae, R. R., & Allik, J. (Eds.). (2002). *The Five-Factor Model of Personality Across Cultures*. Amsterdam: Kluwer Academics.

McNeill, D. (1992). *Hand and Mind — What Gestures Reveal about Thought*. Chicago: The University of Chicago Press.

Miller, C. A., Wu, P., Vakili, V., Ott, T., & Smith, K. (2009). Culture, Politeness and Directive Compliance. In Stephanidis, C. (Ed.), *Universal Access in HCI, Part 1, HCII 2009* (pp. 568–577). Berlin: Springer.

Nakano, Y., & Nishida, T. (2005). Awareness of Perceived World and Conversational Engagement by Conversational Agents. In T. Nishida (Ed.), *Proceedings of the AISB 2005 Symposium on Conversational Informatics for Supporting Social Intelligence & Interaction.*

Nakano, Y., & Rehm, M. (2009). Multimodal Corpus Analysis as a Method to Ensure Cultural Usability of Embodied Conversational Agents. In Kurosu, M. (Ed.), *Human Centered Design, HCII 2009* (pp. 521–530). Berlin: Springer. doi:10.1007/978-3-642-02806-9_60

Nakano, Y., & Reinstein, G. Stocky. T., & Cassell, J. (2003). Towards a Model of Face-to-face Grounding. In *Proceedings of the Association for Computational Linguistics*.

Nazir, A., Lim, M. Y., Aylett, R., Cawsey, A., & Enz, S. (2009). Culture-Personality based Affective Model. *AI & Society*. *Special Issue on Enculturating HCI*, *24*(3), 281–293.

Reeves, B., & Nass, C. (1996). *The Media Equation – How People Treat Computers, Television, and New Media Like Real People and Place*. Cambridge, UK: Cambridge University Press.

Rehm, M. (2008). She is just stupid – Analyzing user-agent interactions in emotional game situations. *Interacting with Computers*, *20*(3), 311–325. doi:10.1016/j.intcom.2008.02.005

Rehm, M., & André, E. (2007). More Than Just a Friendly Phrase: Multimodal Aspects of Polite Behavior in Agents. In Nishida, T. (Ed.), *Conversational Informatics* (pp. 69–84). New York: Wiley. doi:10.1002/9780470512470.ch4

Rehm, M., André, E., Bee, N., Endrass, B., Wissner, M., & Nakano, Y. (2009). Creating Standardized Video Recordings of Multimodal Interactions Across Cultures. In Kipp, M. (Eds.), *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications* (pp. 138–159). Berlin: Springer.

Rehm, M., André, E., Conradi, B., Hammer, S., Iversen, M., & Lösch, E. (2006). Location-Based Interaction With Children For Edutainment. In *Perception and Interactive Technologies* (pp. 197–200). Berlin: Springer. doi:10.1007/11768029_22

Rehm, M., Bee, N., & André, E. (2008). Wave like an Egyptian – Accelerometer based gesture recognition for culture-specific interactions. In Proceedings of Britisch HCI (pp. 13-22).

Rehm, M., Nakano, Y., André, E., & Nishida, T. (2008). Culture-specific first meeting encounters between virtual agents (Honourable Mention). In H. Prendinger, J. Lester & M. Ishizuka (Eds.), *Intelligent Virtual Agents (IVA)* (pp. 223-236). Berlin: Springer.

Rehm, M., & Rosina, P. (2008). Second Life as an Evaluation Platform for Multiagent Systems Featuring Social Interactions. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Rehm, M., Vogt, T., Bee, N., & Wissner, M. (2008). Dancing the Night Away – Controlling a Virtual Karaoke Dancer by Multimodal Expressive Cues. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1249-1252).

Ruttkay, Z. (2009). Cultural Dialects of Real and Synthetic Emotional Facial Expressions. *AI & Society*. *Special Issue on Enculturating HCI*, *24*(3), 307–315.

Schwartz, S. H., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, *26*(1), 92–116. doi:10.1177/0022022195261007

Ting-Toomey, S. (1999). *Communicating Across Cultures*. New York: The Guilford Press. Warren, R., Diller, D. E., Leung, A., Ferguson, W., & Sutton, J. L. (2005). Simulating scenarios for research on culture and cognition using a commercial role-play game. In M. E. Kuhl, N. M. Steiger, F. B. Armstrong & J. A. Joines (Eds.), *Proceedings of the 2005 Winter Simulation Conference*.

KEY TERMS AND DEFINITIONS

Cultural Dimensions: Prominent cultural theory by Hofstede (2001) that defines culture as a concept consisting of five value dimensions. For a concrete culture a value can be given for each dimension. Dimensions are hierarchy, gender, identity, uncertainty avoidance, and orientation.

Cultural Heuristics: Our cultural backgrounds largely depend how we interpret interactions with others, which aspects we find relevant, and what kind of behavior is deemed annoying or insulting. We use the term cultural heuristics to denote such behavioral patterns related to cultural backgrounds.

Embodied Conversational Agent (ECA): Virtual characters serve as communication partners for the user. Apart from verbal interactions, the embodiment allows realizing non-verbal interaction channels like gaze, facial expressions or gestures. Interaction modeling concentrates on the communicative functions of verbal and non-verbal behavior.

Enculturating Interactive Systems: The challenge of integrating cultural aspects of human interaction in an interactive system. Cultural aspects can consist of the different aspects mentioned in this chapter like proxemics, gaze behavior, appearance etc.

Multimodal Corpus: A multimodal corpus is a collection of video data that is annotated along the timeline in order to code information in the video like e.g. gestural expressivity, emotions, or dialogue functions. The annotation serves as an empirical foundation for modeling the behavior of an agent.

Multimodal Interaction: The use of more than one input and output channel is called multimodal interaction, e.g. speech and gestures for input and text and sound for output.

Proxemics: Spatial behavior in face-to-face interactions has been termed proxemics by Hall (1966). He distinguished different spatial areas like personal and social that are linked to different routine behavior.